

Structural Comparison of ColH and ColG Collagen-Binding Domains from *Clostridium histolyticum*

Ryan Bauer,^a Jeffrey J. Wilson,^a Sagaya Theresa Leena Philominathan,^{a*} Dan Davis,^a Osamu Matsushita,^b Joshua Sakon^a

Department of Chemistry and Biochemistry, University of Arkansas, Fayetteville, Arkansas, USA^a; Department of Bacteriology, Okayama University Medical School, Okayama, Japan^b

Clostridium histolyticum secretes collagenases, ColG and ColH, that cause extensive tissue destruction in myonecrosis. The C-terminal collagen-binding domain (CBD) of collagenase is required for insoluble collagen fibril binding and subsequent collagenolysis. The high-resolution crystal structures of ColG-CBD (s3b) and ColH-CBD (s3) are reported in this paper. The new X-ray structure of s3 was solved at 2.0-Å resolution ($R = 17.4\%$; $R_{\text{free}} = 23.3\%$), while the resolution of the previously determined s3b was extended to 1.4 Å ($R = 17.9\%$; $R_{\text{free}} = 21.0\%$). Despite sharing only 30% sequence identity, the molecules resemble one another closely (root mean square deviation [RMSD] $C_{\alpha} = 1.5$ Å). All but one residue, whose side chain chelates with Ca^{2+} , are conserved. The dual Ca^{2+} binding site in s3 is completed by an unconserved aspartate. Differential scanning calorimetric measurements showed that s3 gains thermal stability, comparable to s3b, by binding to Ca^{2+} (*holo* $T_m = 94.1^\circ\text{C}$; *apo* $T_m = 70.2^\circ\text{C}$). *holo* s3 is also stabilized against chemical denaturants urea and guanidine HCl. The three most critical residues for collagen interaction in s3b are conserved in s3. The general shape of the binding pocket is retained by altered loop structures and side chain positions. Small-angle X-ray scattering data revealed that s3 also binds asymmetrically to minicollagen. Besides the calcium-binding sites and the collagen-binding pocket, architecturally important hydrophobic residues and the hydrogen-bonding network around the *cis*-peptide bond are well conserved within the metallopeptidase subfamily M9B. CBDs were previously shown to bind to the extracellular matrix of various tissues. Compactness and extreme stability in physiological Ca^{2+} concentration possibly make both CBDs suitable for targeted growth factor delivery.

Clostridium histolyticum is one of the causative agents for clostridial myonecrosis. The organism produces collagenases responsible for extensive tissue destruction. Although collagenases are harmful during infection, their ability to break down a wide variety of collagen types makes them beneficial as a treatment for excessive connective tissue buildup. The enzymes are approved by the Food and Drug Administration (FDA) to break down the tough cords in Dupuytren's contracture (1). The two classes of collagenase, ColG and ColH, differ in domain structures (s1, s2, s3a, and s3b for ColG and s1, s2a, s2b, and s3 for ColH). The s1 structure is the catalytic domain that belongs to a metallopeptidase M9 subfamily B (M9B) of bacterial collagenases. The structure of ColG s1 was recently solved and revealed an overall saddle shape built up by an activator domain linked to a peptidase domain (2). The amino acid sequences of s2, s2a, and s2b resemble a polycystic kidney disease (PKD) domain. The role of this domain is speculative. It was shown that the PKD-like domain of the serine protease from *Pseudoalteromonas* sp. SM9913 could bind to and swell collagen microfibrils (3). The PKD-like domain has also been shown to enhance collagen binding (4). The C-terminal domains s3a, s3b, and s3 (Swiss-Prot, Q46085; PubMed, BAA34542) are homologues with approximately 120 amino acid residues and are classified as bacterial prepeptidase C-terminal domains (PPC superfamily). The PPC domain is found in some members of metalloprotease families M4, M9, and M28 as well as in a serine protease family, S8. *In silico* work on the PPC domain suggests a distant relationship with the PKD-like domain (5). Moreover, the role of the PPC domain could be diverse. Some of the domains are removed after secretion for activation of certain proteases (5), but others are retained in the mature enzymes (6). Both ColG-derived s3a and s3b and ColH-derived s3 bind to either insoluble collagen *in vitro* or to collagen fibrils in the extracellular matrix of various

tissues (7, 8). Therefore, each of these PPC domains is subclassified as a collagen-binding domain (CBD).

CBD derived from ColG (s3b) has been extensively characterized. It is a β -sandwich fold, and a pocket formed on a β -sheet was identified to interact with the substrates by alanine-scan mutagenesis and minicollagen binding assays (9). The s3b binds to collagenous peptides with triple-helical conformation but not to the similar peptides lacking triple-helical conformation or to gelatin (denatured collagen), suggesting that the CBD-collagen interaction is conformation specific (4, 7). The most recent solution studies of s3b in complex with collagenous peptides by small-angle X-ray scattering (SAXS) and heteronuclear single quantum coherence (HSQC) nuclear magnetic resonance (NMR) titration of various spin-labeled collagenous peptides to ^{15}N -labeled s3b showed that s3b binds unidirectionally to the C terminus of collagenous peptides (10).

In the presence of Ca^{2+} , s3b shows shortened hydrodynamic radius, better stability, and more efficient substrate binding (11). The X-ray crystal structures of s3b were solved in the presence of

Received 5 January 2012 Accepted 2 November 2012

Published ahead of print 9 November 2012

Address correspondence to Joshua Sakon, jsakon@uark.edu.

* Present address: Sagaya Theresa Leena Philominathan, Alexion Pharmaceuticals, Cheshire, Connecticut, USA.

R.B., J.J.W., and S.T.L.P. contributed equally to this article.

Supplemental material for this article may be found at <http://dx.doi.org/10.1128/JB.00010-12>.

Copyright © 2013, American Society for Microbiology. All Rights Reserved.

doi:10.1128/JB.00010-12

Ca^{2+} (*holo*) as well as in the absence of Ca^{2+} (*apo*) to show a secondary structure transformation of the linker at its N terminus (11). The linker is coiled as an α -helix in the *apo* form but is unwound to form a parallel β -strand in the *holo* form, which possibly results in domain rearrangement. Two Ca^{2+} ions bind cooperatively with macroscopic association constants ($K_1 = 5.01 \times 10^5 \text{ M}^{-1}$ and $K_2 = 2.28 \times 10^5 \text{ M}^{-1}$) to trigger the structural transformation (12). Two Ca^{2+} ions chelate to s3b in close proximity to each other (3.7 Å). The N-terminal linker of s3b adopts an energetically unfavorable nonprolyl-*cis*-peptide in the presence of Ca^{2+} . Simulations provided mechanical insights of calcium ions catalyzing the *trans-cis* isomerization of the nonprolyl peptide bond (13). Previously, it was shown that ColG-s1 does not require the PKD-like domain or CBDs to degrade gelatin (4) or soluble collagen (2). These domains, however, are necessary to degrade collagen fibrils.

The s3 from ColH has been shown to anchor growth factors to the extracellular matrix and extend their half-lives. A fusion protein of basic fibroblast growth factor and s3 domain stayed at the site of injection for more than 10 days to promote cell proliferation (14). Although various preclinical applications are currently being examined (14–18), s3 has not been characterized in detail. The s3 and s3b domains share 30% sequence identity. Only three aromatic residues within nine residues in s3b involved in the substrate binding are conserved in s3. Based on the sequence alignment, one of the Ca^{2+} -chelating aspartates (Asp927) in s3b is replaced by serine (Ser896) in s3 (note that in this article, the sequence number for each CBD refers to the mature protein: the s3b sequence numbering reflects a 110-residue-long prepro peptide being cleaved, while s3 numbering reflects a 40-residue-long prepro peptide being cleaved). Each of the two side chain oxygen atoms of Asp927 chelates with a different Ca^{2+} ion. It was uncertain as to how Ser896 in s3 could construct the dual calcium-binding site in place of the bidentate Asp. To address its structure-and-function relationship applicable for various drug delivery applications (14–18), structural work on s3 was initiated. The structures of dissimilar CBD molecules in the metalloproteinase family M9B enabled us to compare and contrast their Ca^{2+} sites and collagen-binding mechanism.

MATERIALS AND METHODS

Expression and purification of collagen-binding domain. Expression and purification of collagen-binding domain as a glutathione S-transferase (GST) fusion protein were achieved using methods described by Matsushita et al. (7).

Crystallization and structure determination of s3 from ColH. Recombinant s3 (7.0 mg/ml in 10 mM Tris-HCl, pH 7.5) crystals were grown by hanging drop vapor diffusion from 28% (wt/vol) PEG3350, 0.1 M Li_2SO_4 , 0.1 M Tris-HCl (pH 6.5), and 2 mM CaCl_2 . The crystals were subsequently flash frozen in liquid nitrogen without the need of additional cryoprotectant and stored in liquid nitrogen until data collection. An in-house X-ray diffraction facility (Rigaku 007, Osmic Blue confocal mirrors, Saturn CCD detector) was used to characterize the unit cell parameters for s3. The calcium-bound s3 crystals belong to the orthorhombic space group $P2_12_12_1$, with unit cell size as follows: a, 62.0; b, 66.2; and c, 96.4 Å. The data were collected at -164°C and diffracted to 2.6-Å resolution. Single-crystal diffraction data were then collected at the 19-ID beam line at the Advanced Photon Source (APS) at Argonne National Laboratories. The crystals diffracted to 2.0-Å resolution. Data were processed using HKL3000. A molecular replacement technique using PHASER on a data set truncated to 3-Å resolution located three molecules in an asymmetric unit (19). The $F_o - F_c$ difference map clearly showed two

TABLE 1 Data collection and refinement statistics

Statistic	<i>holo</i> -s3b	<i>holo</i> -s3
Data collection statistics ^a		
Wavelength (Å)	0.98045	0.97937
Space group	$P2_1$	$P2_12_12_1$
a, b, c (Å)	40.9, 59.2, 48.8	62.0, 64.2, 95.4
beta (°)	100.4	90.0
Unique reflections	44,856	26,207
D_{\min} (Å) ^b	16.61–1.35	28.50–2.00
Completeness (%) ^c	99.8 (100)	99.4 (99.9)
I/ σ I (b)	26.3 (2.6)	16.7 (3.1)
R_{means} (%) ^d	6.4 (65.9)	9.9 (56.6)
Redundancies	4.6 (3.2)	4.0 (4.1)
Refinement statistics		
Resolution (Å)	16.61–1.35	28.50–2.00
R-factor (%)	17.9 (30.9)	17.4 (19.5)
R_{free} (%) 5% of data	21.0 (38.0)	23.3 (28.4)
Avg B-factor main chain in A (Å ²)	11.6	25.2
Avg B-factor side chain in A (Å ²)	14.2	27.5
Avg B-factor main chain B (Å ²)	15.8	21.4
Avg B-factor side chain B (Å ²)	17.8	23.8
Avg B-factor main chain C (Å ²)	NA	23.4
Avg B-factor side chain C (Å ²)	NA	26.2
RMS deviations from restraint target values		
Bond lengths (Å)	0.01	0.02
Angle (°)	1.2	1.7
Distances from restraint planes	0.004	0.007
Ramachandran statistics		
Most favorable (%)	92.1	90.2
Additionally allowed (%)	7.9	9.8
Disallowed (%)	0.0	0.0

^a For more complete data collection statistics, see Tables S7 and S8 in the supplemental material.

^b Highest-resolution shell for *holo*-s3b, 1.40 to 1.35 Å; highest-resolution shell for *holo*-s3, 2.03 to 2.00 Å.

^c Data for the highest-resolution shell are given in parentheses.

^d $R_{\text{means}} = \sum_h \sqrt{\frac{n_h}{n_h - 1} \frac{\sum_i |I_{hi}| - \langle I_h \rangle / \sum_i \langle I_h \rangle}{n_h - 1}}$.

calcium atoms that were omitted from the search model and increased confidence in the phase information. Where possible, amino acid residues of molecule A were corrected to those of s3. Using the partially built molecule A as a search model, molecular replacement techniques were repeated, yielding phases that allowed for building the complete s3. The models were refined using REFMAC to 2.0-Å resolution (20). In the strong difference map peaks ($F_o - F_c$ with σ of >3.5 and $2F_o - F_c$ with σ of >1.0), 326 solvent molecules were built. Noncrystallographic symmetry (NCS) restraints were not used during the refinement. PROCHECK was used to ensure the stereochemical quality of the protein. Data collection and refinement statistics are summarized in Table 1.

Crystallization and structure refinement of s3b from ColG. Recombinant s3b (5.5 mg/ml in 50 mM Tris-HCl, pH 7.5) crystals were grown by hanging drop vapor diffusion from 25 to 27% PEG 3350, 100 mM sodium acetate (pH 4.6), and 600 mM LiCl at 4°C. As it is known to exist in two structural forms, s3b was supplemented with $\text{Ca}(\text{NO}_3)_2$ to a concentration of 50 mM in order to push the equilibrium toward the ion-bound form (11). Crystals were dragged through 100% paraffin oil to remove excess solvent and then frozen in liquid nitrogen. The initial conditions suitable to grow s3b crystals were identified by the sitting drop method using a high-throughput screen (Hampton Research Crystal Screen HT).

TABLE 2 SAXS-derived molecular parameters^a

Mini-collagen: CBD complexes	SAXS		Discrepancy	
	D_{\max} (Å)	R_g (Å) (expt)	R_g (Å) (calc)	χ^2
[(POG) ₁₀] ₃ :S3	87	23.11 ± 0.09	24.43	0.47
[(POG) ₁₀] ₃ :S3b	93	22.62 ± 0.04	24.79	0.75

^a expt, experimentally obtained; calc, calculated.

Cryoprotection procedures resulted in a low mosaic spread, 0.46°, and extended our data to 1.35 Å (unique reflections, 44,856) versus 1.65-Å resolution (unique reflections, 27,154) for our previously deposited structure (11). The data were collected at BioCARS 14ID-B at Advanced Photon Source. All the data were indexed and scaled using Denzo and Scalepack from HKL2000 (21). SHELX97 (22) and REFMAC (20) were used for refinement (Table 1). The *holo* structure (Protein Data Bank [PDB] code 1NQD) was used as a starting model. The manual adjustments were aided by the use of MIFit (23). Seven alternate conformations for surface residues (919, 959, 968, and 976 of molecule A and 895, 968, and 976 of molecule B) were built. The positional and anisotropic B-factor refinements were carried out using Shelxl97 initially (22) and then REFMAC (20). Hydrogen atom positions were constrained to the ideal position during the refinement. However, neither noncrystallographic symmetry restraints nor Ca-O bond restraints were applied. Positional refinement, manual adjustments, and anisotropic B-factor refinement resulted in an *R* of 17.9% and an *R*_{free} of 21.0% (Table 1). The anisotropic temperature refinement is justified because *R*_{free} dropped nearly 5%. Seventy percent of all atoms, including water molecules, are very well ordered (B-factor below 30 Å²), and their positional errors are less than 0.05 Å.

SAXS experiments. The SAXS data were collected on solutions of s3, various collagenous peptides, and the CBD-collagenous peptide complexes (between 4 and 6 mg/ml) in 10 mM Tris-HCl (pH 7.5) containing 100 mM NaCl and 20 mM CaCl₂ at the XOR beamline of sector 12-ID at the Advanced Photon Source in the Argonne National Laboratory. The main advantage of X-ray scattering is that it can be carried out in solution under nearly physiological conditions (13). The monochromatic radiation source (10 keV, $\lambda = 1.2398$ Å) was the APS Undulator A insertion device using a Si-111 monochromator, with 1:1 horizontal focusing and higher harmonic rejection from an Rh-coated mirror, and beam-defining slits set at 0.3 mm (vertical) and 0.25 mm (horizontal). A 1.6-mm diameter capillary flow cell with a flow rate of 4 μl/second was used to capture four frames with exposure time of 10 s. The SAXS detector used was a Mar165 scintillator fiber optic-coupled charge-coupled-device (CCD) detector and covered the momentum transfer range of $0.005 < q < 0.198$ Å⁻¹, where q is $4\pi \sin\theta/\lambda$ (2θ is the scattering angle). The WAXS detector was a custom-made Roper scintillator fiber-optic-coupled CCD detector and covered the range of $0.191 < q < 1.8$ Å⁻¹ S (24).

All scattering data were acquired at 10°C. The four scattering patterns from each detector were averaged and merged with the rejection of outlying scans. For further analysis, the program IGOR Pro 5.5A (WaveMetrics) was used. The scattering profiles of the protein, peptide, and their complexes were obtained after subtracting the buffer profiles. The reduced scattering data were plotted as scattering intensity $I(Q)$ versus Q (see Fig. S1A in the supplemental material). The radius of gyration, R_g , was obtained from the Guinier approximation by linear least-squares fitting in the $QR_g < 1$ region, where the forward scattering intensity, $I(0)$, is proportional to the molecular weight of the protein complex. An indirect Fourier transformation of the $I(Q)$ data, using GNOM (25), provided the pair-distance distribution function $P(r)$ in the real space. The point at which $P(r)$ intersects with the x axis represents the maximum diameter, D_{\max} (Table 2; see also Fig. S1B in the supplemental material). The molecular envelopes were constructed from the SAXS data after *ab initio* calculation with the program GASBOR (26). The $P(r)$ ranges were varied to generate various (>10) molecular envelopes using GASBOR (26). No

symmetry restraints were applied to any of the shape reconstructions. The 10 highest-scoring *ab initio* models were chosen and averaged using DAMAVER (27). Atomic models were docked into *ab initio* envelopes with the program SUBCOMB (28). The theoretical scattering curves from the atomic models were calculated and compared with the experimental curves with the program CRY SOL (29). Rigid-body modeling to the experimental scattering data was performed using SASREF (30).

Docking model. The mini-collagen: CBD complex was generated from the CBD and 1K6F for the collagenous peptide. To obtain the complex, the soft docking algorithm BiGGER was used (31). The manual adjustments were aided by the use of MIFit (23).

Fluorescence spectroscopy measured equilibrium denaturation of s3. For s3, λ_{\max} for the native protein occurs at 310 nm while λ_{\max} for the denatured protein occurs at 350 nm. The ratio of intensity at 350 nm versus the intensity at 310 nm was used to track the unfolding process. Fluorescence data were collected on a Hitachi F-2500 fluorimeter with excitation and emission bandwidths at 2.5 nm and 10 nm, respectively. The excitation wavelength used was 280 nm, and fluorescence emissions were monitored between 300 nm and 450 nm. During thermal denaturation trials, the temperature of the protein solution was maintained with a Neslab RTE-110 circulating water bath (Thermo Scientific, Newington, NH). In the thermal denaturation trials, the protein concentration was 3 μM. In the chemical denaturation trials, the protein concentration was 1.5 μM. *Holo*-s3 was supplemented with 1 mM CaCl₂, while *apo*-s3 was supplemented with 1 mM EDTA. In all cases, the protein was diluted in 10 mM Tris-HCl and 100 mM NaCl, and the pH was maintained at 7.5. When urea was used as the denaturant, the native s3 was exposed to concentrations of denaturant that increased linearly by 0.2 M intervals from 0.0 to 9.8 M. When guanidine hydrochloride (GuHCl) was used as the denaturant, s3 was exposed to concentrations of denaturant that increased linearly by 0.2 M intervals from 0.0 to 5.8 M. During thermal denaturation, s3 was exposed to temperatures that linearly elevated by 2.5°C intervals from 7.5 to 100°C. ΔG_{H_2O} and *m* values were calculated as described previously (12).

DSC. Differential scanning calorimetry (DSC) was conducted using a nanodifferential scanning calorimeter (NDSC) model CSC 6300 (Calorimetry Sciences Corporation). The s3 was dissolved in 10 mM Tris-HCl (pH 7.5) and 1 mM CaCl₂ for the *holo* form to a final concentration of 0.5 mg/ml. The *apo*-s3 was dissolved in 10 mM Tris-HCl (pH 7.5) to a final concentration of 0.5 mg/ml. Following a scan of buffer versus buffer to establish the baseline, samples were run as follows. Degassed samples were heated from 25°C to 130°C (1°C/min) and then cooled back to the initial temperature at the same rate. A second heating scan was then performed to test the reversibility of the process. Unfolding was reversible if a difference in heat capacity was again observed during the second heating scan. DSC analysis software (CpCalc1) provided by the instrument manufacturer was used to analyze the transition temperature after subtracting the baseline.

Protein structure accession numbers. The full coordinates of the structure of *holo* s3b and *holo* s3 and their corresponding structure factor amplitudes have been deposited in Protein Data Bank under accession codes 4HPK and 3JQW. s3 solved in the presence of calcium/cadmium is discussed in the supplemental material (PDB accession code 3JQX).

RESULTS AND DISCUSSION

Structure description of s3 at 2.0-Å resolution. The structure of s3 from ColH is similar to the structure of s3b from ColG. The s3 also adopts a β-sandwich “jelly roll” conformation composed of nine β-strands. Two calcium atoms per molecule are found between the linker (861 to 878) and the β-strand C. There are three molecules in the asymmetric unit. Molecule A consists of 117 amino acids (863 to 979), molecule B consists of 120 amino acids (862 to 981), and molecule C consists of 121 amino acids (861 to 981). The backbone structures of the three molecules (A to C) are virtually identical. The root mean square deviations of C_α posi-

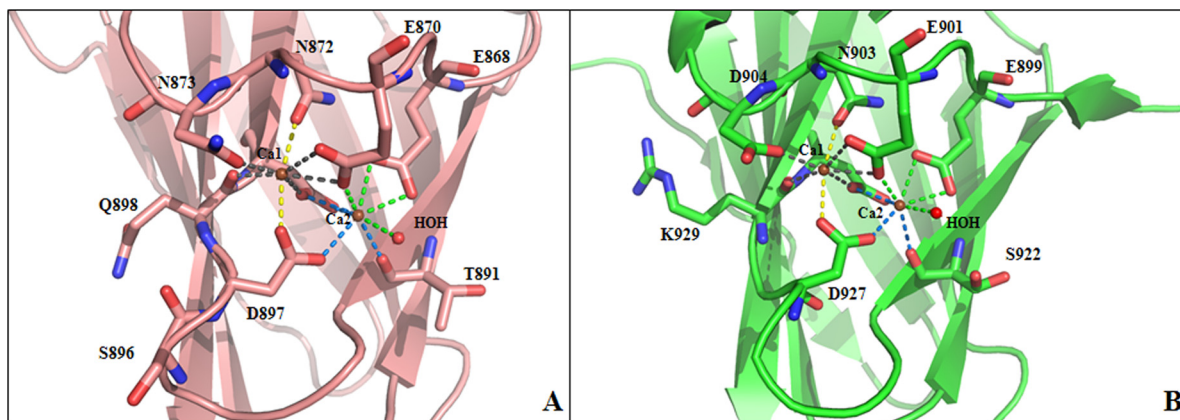


FIG 1 Ca^{2+} coordination in s3 and s3b. Comparison of Ca^{2+} coordination of s3 (A) and s3b (B). For both CBDs, the pentagonal base around Ca1 is indicated by gray dashes, while the axial positions are indicated by yellow dashes. Seven oxygen atoms from six residues interact with this Ca^{2+} . For Ca2 of both CBDs, one square face is indicated by blue dashes, while the other face is indicated by green dashes. Seven oxygen atoms from five residues and a water molecule interact with this Ca^{2+} . Each oxygen atom is labeled with residue type and residue number. D897 of s3 fulfills the role of D927 of s3b. This figure was prepared using PyMOL (48).

tions between molecules A and B, determined using Superpose from the CCP4 suite of programs, was 0.29 Å, that between B and C was 0.45 Å, and that between A and C was 0.43 Å. Three molecules were independently refined without NCS restraints. Inherent flexibility was observed at both the N terminus (amino acids 861 to 863) and the C terminus (amino acids 977 to 981). Another deviation occurs only in molecule C at the loop between β -strands H and I (amino acids 966 to 969) caused by a peptide flip between Ser967 and Tyr968. Both calcium atoms are ordered (B-factor, $<18 \text{ Å}^2$).

Structural comparison of holo s3b and holo s3. Here we present the first structure of a collagen-binding domain from ColH collagenase. The tertiary structures of the CBD molecules are very similar to one another (C_α RMSD = 1.4 Å), despite being of low sequence identity. β -Sheet regions are the most similar. Even the linker residues 898 to 910 of s3b and 867 to 879 of s3 adopt very similar structures. The Ser874-Thr877 segment of s3 adopts a short 3_{10} helix, as also observed in the Ser905-Lys908 segment of s3b. In both structures, two calcium ions bind between the N-terminal linker and the β -strand C. The s3 does not adopt β -strand A found in s3b, and the region adopts a random coil. Also, the solution structure of apo-s3b does not form β -strand A (32). The crystal structures of s3 and s3b are most divergent at the N terminus and the loops connecting the β -strands. For example, a loop of 8 residues (Glu961-Ile968) between β -strands E and F in s3b are line broadened and nonassignable on NMR, and the flexible loop is unobserved in the crystal structure (10, 11). In contrast, the equivalent loop in s3 (Glu930-Val935) is well ordered. Another significant deviation in loop structures occurred at a loop between strands B and C (residues 924 to 928 in s3b and 894 to 899 in s3). The N termini of s3b (residues 891 to 897) and s3 (residues 861 to 868) adopt very different paths. It is possible that this deviation can be explained by crystal packing restraining the position of the termini. Terminus deviation as a result of crystal packing, for example, was seen in the N-terminal linker of apo-s3b (11, 32).

Coordinations around the two calcium atoms in s3 are virtually identical to those described in detail for s3b later. Seven oxygen atoms of five side chains and one main chain form a pentagonal-bipyramid coordination (Fig. 1). Eight oxygen atoms of four

side chains and a main chain and one water molecule form a square antiprismatic coordination. The average bond distances observed in s3 are similar to those in s3b (see Tables S2 and S3 in the supplemental material). There are some differences between the two CBD molecules. Asp904 in s3b is replaced by Asn873 in s3; therefore, the net charge around the calcium sites in s3 is -4 . An earlier sequence alignment of CBD molecules showed that Asp927 in s3b is replaced by Ser896 in s3 (11). However, serine is rarely found to chelate with Ca^{2+} (33). The crystal structure of s3 revealed that Asp897 is found in place to chelate with Ca^{2+} atoms and Ser896 is pointed away from the ions. In order for Asp897 to chelate equivalently to Asp927 of s3b, a loop (residues 894 to 898) of s3 meanders differently to position OD1 and OD2 atoms of Asp897. Two energetically unfavorable torsional strains were found in s3b (11). One of them is the peptide bond between residues Glu901 and Asn902 in s3b, which adopts an uncommon and energetically unfavorable nonprolyl-*cis*-peptide conformation. OD1 of Asn902 makes a hydrogen bond with main-chain N of Asp904. The interaction is one of the important hydrogen bonds in the stabilization of the transition state for the peptide isomerization (13). In contrast, the structurally equivalent peptide bond between Glu870 and Pro871 in s3 adopts a common prolyl-*cis*-peptide conformation (Fig. 2). The second torsional strain was found in the side chain χ_1 angle of residue Tyr931 of s3b. Tyr931 adopts an energetically unfavorable angle of 64° , which is stabilized by hydrogen bonding with NE2 of His959 of s3b (distance, 2.6 Å). The equivalent residue Tyr900 of s3 adopts a favorable χ_1 angle of -68° (Fig. 3). The s3b domain is extremely thermostable ($T_m = 93^\circ\text{C}$) despite the torsional strains (12). The s3 domain is only slightly more stable ($T_m = 94^\circ\text{C}$). The stability of s3 is described in more detail in the following section.

The collagen-binding surface in s3b was identified by binding assay of 32 surface mutants (11) and by HSQC-NMR titration (10). Three aromatic side chains critically involved in collagen binding are positioned similarly in s3. Two tyrosine residues (Tyr970 and Tyr994) in s3b are conserved in s3 (Tyr937 and Tyr962), but Tyr996 is replaced by phenylalanine (Phe964) in s3. Seven other residues identified to interact with the collagenous peptides are not conserved in s3. SAXS and NMR studies of s3b in

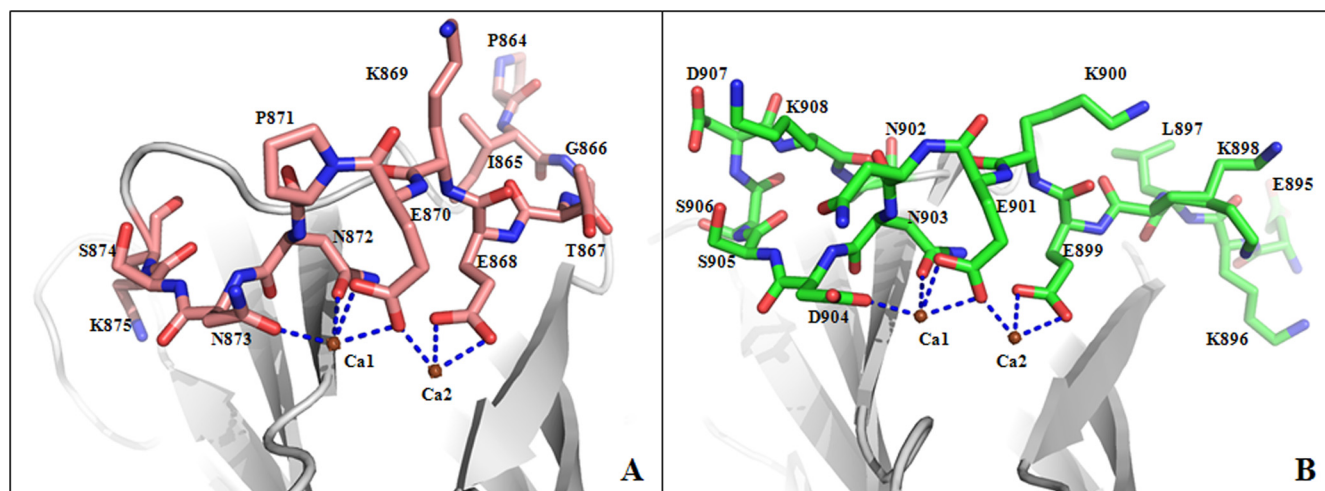


FIG 2 Linker of CBD. (A) Close-up view of linker in s3 (ColH). The peptide bond between Glu870 and Pro871 adopts prolyl *cis* conformation. Glu868, Glu870, Asn872, and Asn873 in the linker region chelate with Ca^{2+} ions (brown spheres). (B) Close-up view of the linker in s3b (ColG). The peptide bond between Glu901 and Asn902 adopts an energetically unfavorable nonprolyl *cis* conformation. Glu899, Glu901, Asn903, and Asp904 in the linker region chelate with Ca^{2+} atoms (brown spheres). This figure was prepared using PyMOL (48).

complex with collagenous peptides showed that Ser928 and Arg929 interact with the C terminus of collagenous peptide (10). Neither residue is conserved in s3. In s3, main-chain carbonyl Thr895 occupies the space where the Ser928 side chain is found, and the three-dimensional space occupied by Arg929 in s3b is occupied by Gln898 and Tyr900 in s3 (see Fig. S2 in the supplemental material). Despite substitutions, the resulting collagen-binding pocket in s3 closely resembles that of s3b. Using SAXS of s3 in complex with $[\text{G}(\text{POG})_7\text{PRG}]_3$ showed that s3 also binds asymmetrically to the collagenous peptide (Fig. 4). Our previous work demonstrated that both $(\text{POG})_{10}$ (L-hydroxyproline is represented as O) and $\text{G}(\text{POG})_7\text{PRG}$ adopt a triple-helical structure in solution. They are thus represented as $[(\text{POG})_{10}]_3$ and $[\text{G}(\text{POG})_7\text{PRG}]_3$, respectively. The resulting atomic models for s3

and s3b fit very well into the *ab initio* SAXS envelopes. The theoretical R_g values calculated from the complex models of s3 and s3b are in excellent agreement with the experimental values (Table 2). The theoretical scattering curves calculated from the models using CRY SOL (29) are superimposable with the experimental curves with minimal discrepancy (Table 2). Rigid body modeling to the experimental scattering data was performed using SASREF (30). Collagenous complexes with either s3 or s3b were fitted independently to the X-ray scattering data. The simulations were repeated 10 times. The model with the lowest value is shown (Fig. 4; see Fig. S3 in the supplemental material). S3b binds asymmetrically to the C-terminal $(\text{POG})_3$ repeats or C-terminal $(\text{POG})_2\text{PRG}$ (10). Since critical aromatic residues are oriented similarly in both structures, it is likely that s3 also targets the C-terminal region. The respective

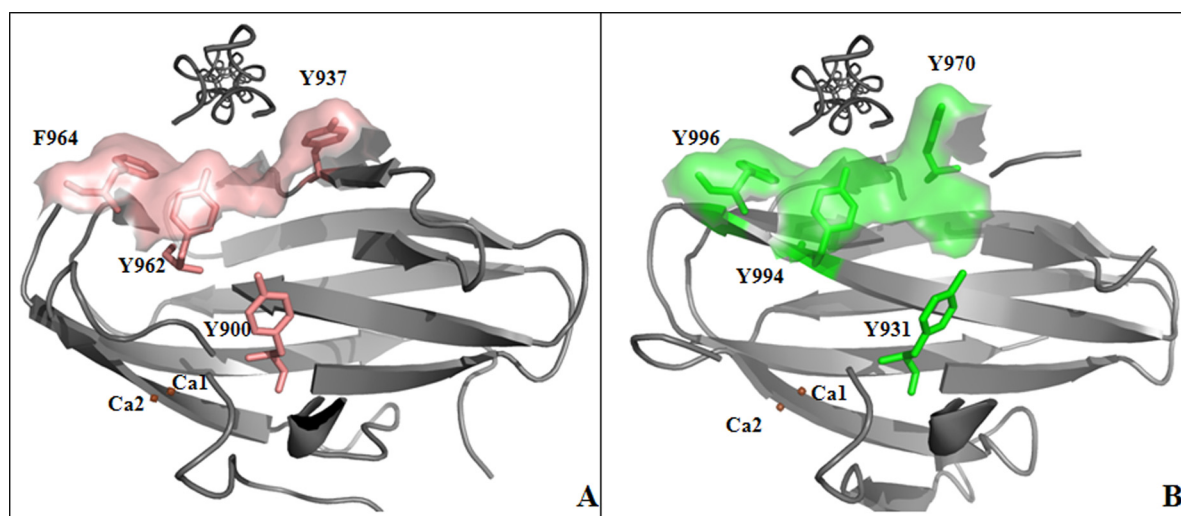


FIG 3 Collagen-binding cleft of CBD. (A) Collagen-binding cleft in s3 (ColH). Conserved residues Thr924, Val926, Tyr937, Tyr962, and Phe964 may shape the binding cleft. Tyr900 adopts a favorable χ_1 , and it is facing the binding cleft. (B) Collagen-binding cleft in s3b. Residues Ser928, Arg929, Thr957, Tyr970, Tyr994, and Tyr996 are the key residues shaping the binding cleft. The collagen-binding cleft matches the width of collagen (~ 10 Å). Tyr931 adopts an atypical χ_1 angle. This figure was prepared using PyMOL (48).

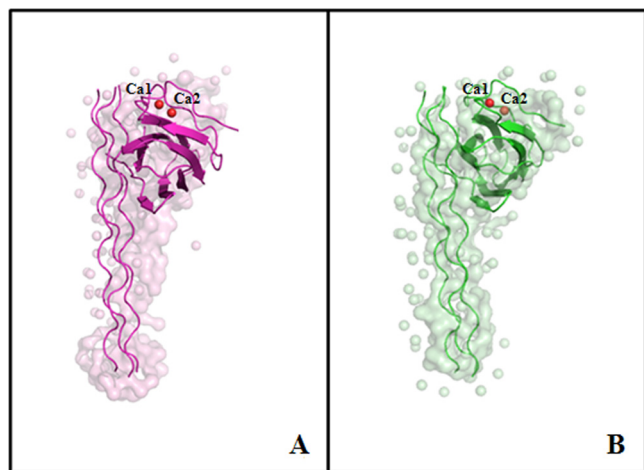


FIG 4 Structures derived from SAXS profiles using *ab initio* simulated annealing calculations are shown as gray surfaces for s3-[G(POG)₈]₃ complex (A) and s3b-[G(POG)₇PRG]₃ complex (B). Calcium ions are shown in red. This figure was prepared using PyMOL (48).

collagenase could target common structural features in collagen fibril. The C-terminal region of collagenous peptide is slightly untwisted (undertwisted). Our studies on s3b and s3 imply that the cleft-like shape of the binding pocket scans the collagen for undertwisted regions and conserved aromatic residues of the pocket intercalate to the triple-helical collagen (34).

Ca²⁺ coordination of s3b at 1.4-Å resolution. The 1.4-Å resolution structure of the Ca²⁺-bound s3b showed excellent electron density for nearly all protein atoms and water molecules. The protein is crystallized as a noncrystallographic symmetry (NCS)-related dimer. Since the structural features have already been described for s3b, the discussion will focus on calcium binding sites. Calcium associates with s3b in two coordination geometries, pentagonal bipyramidal and square antiprismatic (Fig. 1). Both calcium atoms were highly ordered (B-factor < 14 Å²). The pentagonal bipyramidal coordination is commonly observed in calcium chelating proteins, of which calmodulin's EF hand motif is the classic example. Five ligating atoms encircle the calcium ion as a near-planar pentagon, and two remaining ligands sit at axial positions above and below. As in the ligation pattern of EF hands, both carbonyl oxygen atoms of s3b's Glu901 provide two of the five equatorial ligation positions. The atoms describing the equatorial plane have an average deviation from planarity of 0.16 Å (see Table S2 in the supplemental material), which compares favorably to the 0.18-Å average deviations across calcium sites in calmodulin (PDB code 1OSA) (35). While calmodulin has a water molecule occupying one axial position, allowing for greater flexibility in the ligand position, both axial positions in s3b are supplied by amino acid side chains (Asn903 and Asp927).

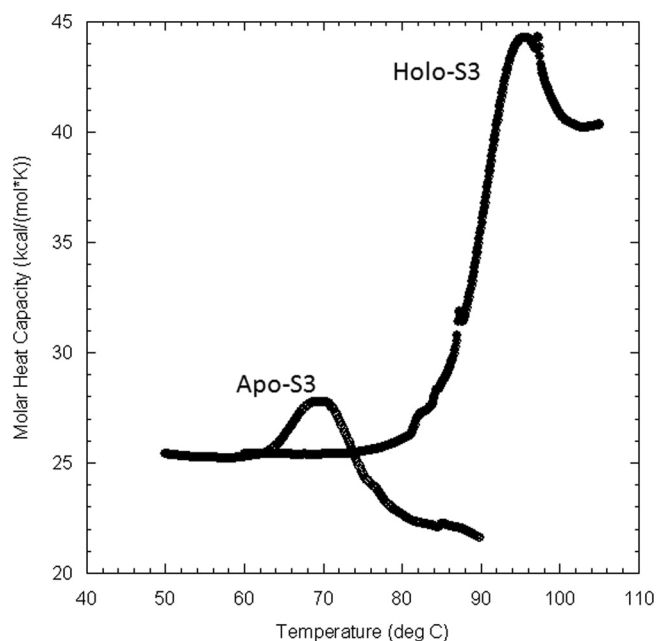
The s3b's second calcium site with square antiprismatic coordination sits 3.7 Å from the pentagonal bipyramidal site. The planar faces of calcium-bound s3b's square antiprism deviate only slightly from a flat geometry (see Table S3 in the supplemental material). One face shows an average out-of-plane deviation by a mere 0.03 Å for coordinated atoms, while the other face shows only 0.13-Å deviations. These values vary from the few other available protein structures with square antiprismatic bound ions (see Table S4 in the supplemental material). Square antiprismatic co-

ordination is relatively uncommon among macromolecules (33), and few are currently found in the Protein Data Bank. Ding et al. described a Zn²⁺-bound RNase T₁, at 1.8-Å resolution, in which six water molecules and a sole aspartate coordinate zinc (36). When defined as square antiprismatic, the carbonyl oxygen atoms of aspartate provide a ligand for each square face, creating significant distortions from planar geometry for these faces and a very short ion-to-plane distance. On further examination, the ion coordination in RNase T₁ is best described as trigonal dodecahedral rather than square antiprismatic, which explains the high deviations from planarity. Potassium ions are found in square antiprismatic coordination in two structures, the KcsA potassium channel and an *Oxytricha nova* G-quadruplex (37, 38). Main-chain carbonyl oxygens coordinate the KcsA potassium ions within the channel, and due to the 4-fold crystallographic symmetry, the square faces are perfectly flat. In the G-quadruplex structure, an unusual DNA structure found in telomeric DNA, potassium ions sit similarly in the center of a channel created this time by nucleotides. Thymine base O6 atoms coordinate the potassium and provide near-planar square faces, with two cubic coordinations sandwiching a square antiprismatic coordination (see Table S4 in the supplemental material).

In addition to the CBDs, two calcium-binding proteins that utilize square antiprismatic geometry are found in the literature. Thermitase, an extracellular serine proteinase, when crystallized in 100 mM calcium, revealed a square antiprismatic calcium site in addition to two pentagonal bipyramidal sites. At lower calcium concentrations, the site exhibits pentagonal bipyramidal coordination (39). Second, alpha-amylase binds two calcium ions with a coordination number (CN) of 8 (40). One of these sites follows a square antiprismatic geometry with three water molecules acting as ligands, and the deviations from planarity are near that of the thermitase coordination (see Table S4 in the supplemental material). However, not all binding sites with CN of 8 have square antiprismatic geometry. The second site in alpha-amylase involves five water molecules and is best described as cubic. Proteinase K binds to one calcium ion with CN of 8 (41), but coordination is dodecahedral in geometry. Four water molecules act as ligands.

The bidentate ligating nature of aspartates and glutamates, often found in ion coordination, distorts geometries from ideal, inorganic values. Therefore, square faces are rarely square, with the exception of proteins taking advantage of multimerization as in KcsA. Previous publications showed that Ca²⁺ is involved in activation but K⁺ is not. However, Li⁺ was found in the Ca²⁺ binding site (11). Barring occupancy differences, Ca²⁺ and K⁺ would have identical electron density in an X-ray diffraction experiment. For the proteins shown in Table S4 in the supplemental material, the distance of calcium from the square faces formed in antiprisms averages 1.37 Å, while potassium distances are a longer 1.58 Å. Additionally, the average ion-to-ligand distance for calcium and potassium are 2.54 Å and 2.81 Å, respectively. The larger coordination radius observed for potassium is not surprising, since its ionic radius is greater than that of calcium. Based on the coordinate errors for oxygen atoms, the errors associated with the bond distances in this study (see Table S4 in the supplemental material) are estimated to be less than 0.03 Å, even accounting for 10% underestimation (9). Based on the analyses of bond distances, it is clear that two Ca²⁺ atoms bind to s3b.

Another novel aspect of the dual calcium site is that it contains the rare tridentate acidic side chains (Glu901 and Asp930). As far

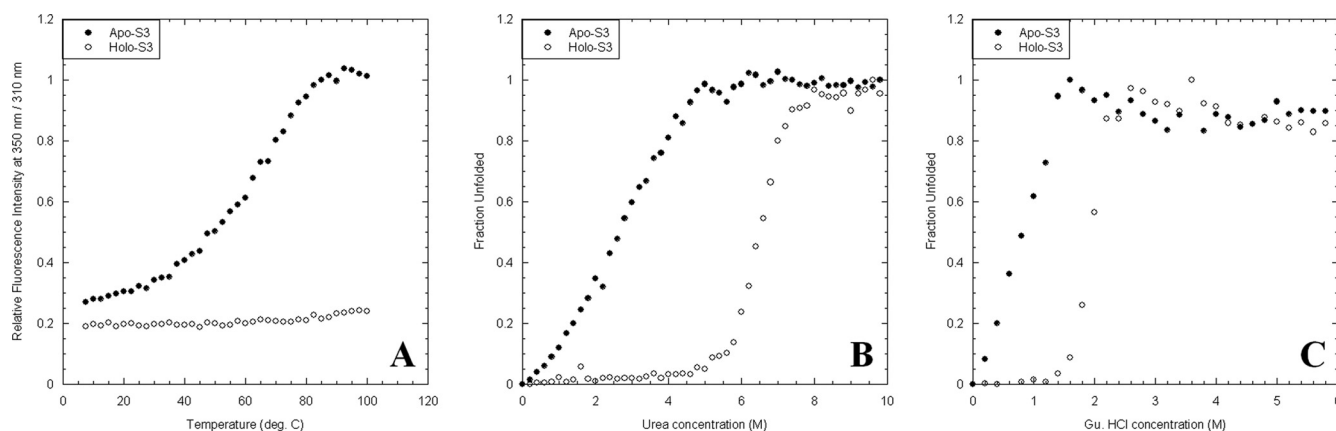
FIG 5 DSC profiles for *apo-s3* and *holo-s3*.

as we are aware, a tridentate acidic side chain has never been observed. Previous analysis of 1,605 Ca^{2+} binding sites showed that bidentate oxygen atoms chelate with longer Ca^{2+} -O distances than those found in monodentate (2.6 Å versus 2.4 Å) (33). Indeed, Ca^{2+} -O in bidentate residues (Glu899 & Asp927) is longer than that in monodentate residues (Asn903 and Asp904). We also note that the two bidentate residues chelate to Ca^{2+} differently. Two carbonyl oxygen atoms of Glu899 side chain chelate with one Ca^{2+} atom, whereas two carbonyl oxygen atoms of Asp927 chelate with two different Ca^{2+} atoms. Ca-O-C angles in Glu899 (89° and 94°) are similar to the values observed for bidentate glutamates in other Ca^{2+} binding proteins ($93.6 \pm 11.3^\circ$) (33), but the angles observed for Asp927 (131° and 144°) are closer to angles observed in monodentate residues ($140.4 \pm 15.2^\circ$) (33). In both tridentate residues (Glu901 and Asp930), one of the three bond distances is significantly shorter (<2.36 Å) than the other two. Ca-O-C angles (87°, 96°, and 159°) are virtually identical for both tridentate res-

idues. Electrostatic repulsion between the cations is likely neutralized by the acidic residues. The coordination geometry reported here could be useful in predicting proteins that utilize acidic residues to chelate a cluster of divalent cations such as Ca^{2+} , Mn^{2+} , and Mg^{2+} .

Stability contributions of calcium binding. Both s3b and s3 are compact and extremely stable in the presence of physiological Ca^{2+} . This may prolong enzyme activity against insoluble collagen in the extracellular matrix. Calcium-bound *holo-s3b* was previously shown to be more stable than *apo-s3b* based on DSC and fluorescence spectroscopy-monitored equilibrium denaturation data (11, 12). Similar to the observation in s3b, denaturation pathways obtained using either DSC (Fig. 5) or fluorescence (Fig. 6) showed that *holo-s3* is more stable than *apo-s3* (see Table S5 in the supplemental material). Stability data were compared with those of s3b. DSC data showed that *holo-s3* ($T_m = 94^\circ\text{C}$) was slightly more stable than *holo-s3b* ($T_m = 93^\circ\text{C}$), though *apo-s3* ($T_m = 70^\circ\text{C}$) was less stable than *apo-s3b* ($T_m = 74^\circ\text{C}$) (see Table S5 in the supplemental material). Whether heat, urea, or guanidine hydrochloride (GuHCl) was used to denature the protein, the same trends were observed. As a side note, the heat denaturation was reversible for *apo-s3* but not for *holo-s3*. Based on fluorescence-monitored thermal denaturation, neither *holo*-CBD fully unfolded at 100°C . Unfolding of s3 was monitored using the fluorescence of intrinsic Trp925, which is sandwiched between β -sheets. Whereas denaturation using urea or GuHCl on *holo-s3* occurred as a cooperative transition from the native to the denatured protein, the unfolding transition of *apo-s3* is abrupt (Fig. 6). *Apo-s3* begins to unfold even at initial levels of chemical denaturant or heat. Meanwhile, the T_m value for *apo-s3* based on fluorescence data is significantly lower than that determined by DSC (see Table S5 in the supplemental material). The discrepancy in T_m was also observed in s3b (12). Analysis of *holo-s3* structure using NACCESS (42) demonstrated that Trp925 is solvent inaccessible. Also, crystal structures of *holo-s3b* exhibited tighter packing around the Trp than did those of *apo-s3b* (12). The folding studies suggest that Trp in *apo-s3* is much less tightly packed than *apo-s3b*.

As observed for s3b, a locally stable, solvent-inaccessible core around Trp925 in *holo-s3* could explain the higher T_m value observed using fluorescence data, while a locally unstable, solvent-

FIG 6 Results of fluorescence-measured equilibrium denaturation of s3 in its *apo* form (closed circles) and in its *holo* form (open circles). (A) Heat denaturation pathway of s3. (B) Urea denaturation pathway of s3. (C) GuHCl denaturation pathway of s3.

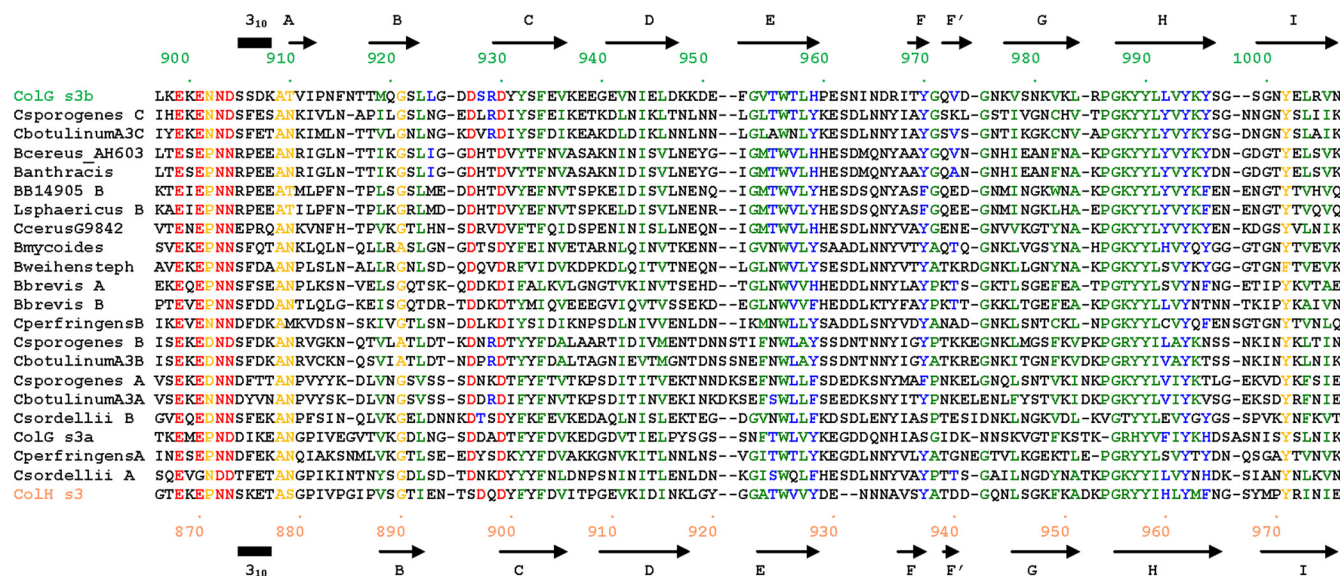


FIG 7 Structure-based sequence alignment of collagen-binding domains from M9B family. Calcium chelating residues, those critical for *cis-trans* linker isomerization, collagen-binding residues, and architecturally important residues are highlighted in red, yellow, blue, and green, respectively. Sequence alignment was aided by the use of ClustalW2 (49).

accessible core around Trp925 in *apo-s3* could explain the lower T_m value observed using fluorescence data.

Evolutionarily related CBD. Both s3b and s3 molecules reported in this paper belong to the PPC superfamily and are collagen-binding domains (CBD). Common structural features described in the previous sections enabled us to update the sequence alignment of the CBD in the M9B subfamily (Fig. 7). Conserved residues are important for one of four reasons: calcium chelation (shown in red), *cis-trans* isomerization of the linker (yellow), collagen-binding (blue), or protein folding (green). Conserved residues in unsolved CBD structures will likely fulfill roles found in s3 and s3b. The dual calcium-binding site is formed by four chelating residues (Glu899, Glu901, Asn903, and Asp904) within the N-terminal linker and two chelating residues (Asp927 and Asp930) from the β -strand C, and invariant Tyr1002 hydrogen bonds and orients Asp930. The residue numbers used in this paragraph are those of s3b. Likewise, other supporting cast such as Gly921 is conserved in the middle of the β -strand, strategically placed to make room for Glu899. The dual calcium chelation site is fashioned sometimes by a neighboring residue. As mentioned, Asp897 of s3 acts equivalently to Asp927 of s3b. Asp897 equivalents are tentatively identified in *Bacillus brevis* s3a and s3b, *Clostridium botulinum* A3 s3a, and *Clostridium histolyticum* ColG s3a. Tridentate and divalent Asp and Glu residues are conserved with only *Clostridium sordellii* s3a as the exception. The monodentate Asp904 residue is sometimes replaced by Asn. For those substituted, the net charge of the dual calcium site is neutral rather than -1 .

A *cis*-peptide bond forms between residues 901 to 902 of *holo-s3b* and 870 to 871 of *holo-s3* (Fig. 2). The residue equivalent to Asn902 in other CBD molecules is either Pro, Asp, or Asn. Pro frequently succeeds the peptide bond to ease *trans-cis* isomerization. In s3b, Asn902_OD1 and Asp904_N form a critical hydrogen bond for the peptide isomerization (13). For CBD molecules where Asp succeeds the *cis* bond, side chain oxygen could play the

same role as Asn902_OD1. Other hydrogen bonds identified by simulation studies important in stabilizing the transition states (13) are well conserved. The simulation study did not account for hydrogen bonding via water-mediated interactions involving Lys900_O, Asn902_O, Lys908_O, and Thr910_OG1, which are also conserved (see Table S6 in the supplemental material). Calcium ions could catalyze the isomerization in all the CBD molecules, and their transition states and catalytic mechanism may look very similar.

Conserved residues are important in ensuring either proper folding or architectural stability. Hydrophobic residues packed between the β -sheets are better conserved if they are located in the vicinity of functionally critical residues. For example, invariant Trp956 of strand E is packed between the β -sheets. The flanking residues (Thr955 and Thr957) interact with minicollagen. Tyr932 is packed between the sheets and helps the positioning of Tyr1002. Residues at tight turns are conserved as well. Gly975 is well conserved to allow a type II' turn in s3b (11). Gly942 (Gly975 equivalent) in s3 allows Asp941 side chain to stabilize the reverse turn. A highly conserved six-residue stretch, 986 to 991 (PGKYLL), adopts a tight turn and precedes the functionally important strand H. The region is well ordered in the crystal structures with low B-factors and is the least dynamic based on NMR (10) and limited proteolysis matrix-assisted laser desorption ionization–time of flight mass spectrometry (MALDI-TOF MS) (32). The main-chain carbonyl and amino groups of Arg985 hydrogen bond with OH of Tyr989 to stabilize the turn. Only Gly987 can make room for the bulky Tyr989 side chain. Tyr990 packs against the invariant Ala909 and conserved 3₁₀ helix. Ala909 is at the base of the linker that undergoes α -helix- β -strand transformation. The tight turn may ensure that collagen-interacting Leu992, Tyr994, and Tyr996 would be correctly positioned. Tyr994 is the most critical residue in interacting with collagenous peptides (11). The strands adjacent to strand H, i.e., strands C and E, are very well conserved. The three antiparallel strands mold the collagen-binding pocket.

Strand F staples the β -sheets by interacting with both sheets. The β -strand first interacts in an antiparallel orientation with strand E and then breaks its direction at Gly971 to interact with strand G. In place of Gly971, Ala or Pro is found at the location where the strand switches its allegiance. The dual interaction of the strand helps position Tyr970 to interact with minicollagen.

Three residues (Tyr970, Tyr994, and Tyr996) shown to interact strongly with minicollagen (10, 11) are conserved. Tyr996 of s3b is a critical residue in binding minicollagen and is partially conserved. This residue is replaced with Phe in s3, though the two side chains have identical orientations. In other CBD molecules, an aromatic residue, such as Phe or His, is sometimes found at the site. Meanwhile, either a β -branched residue or Leu is found at the position equivalent to Thr957 in most of the CBDs. Six other residues identified by ^{15}N -HSQC-NMR titration to interact with minicollagen are not well conserved. Since divergent CBDs (s3 and s3b) adopted a similar cleft-shaped binding pocket, other CBDs may also adopt a similar collagen-binding structure.

Divergent CBD could target different collagen sequences and could possibly target different collagen types; however, this structural study suggests otherwise. Rather, all the CBD domains may bind similarly to an undertwisted region such as the C terminus of a collagen fibril. The C terminus of type I collagen is exposed in the fibril surface based on X-ray fiber diffraction experiments (43), and it is the most accessible site for the bacterial collagenase to initiate assaults. However, CBD binding only at the C-terminal region of tropocollagen is unfounded. Gold particle-labeled tandem ColG-CBD (s3a-s3b) labeled with gold particle bound to type I collagen fibrils exhibited no periodicity (8). In the collagen fibrils, the molecules are staggered from each other by about 67 nm. Therefore, CBD could target partially undertwisted regions in the middle of a tropocollagen that are also vulnerable for assaults (34). Both type I and II tropocollagen molecules are suggested to consist of a triple helix with some relaxations (44). The energetic requirement to unwind these regions could be much less than for tightly wound helical regions. Both s3 and s3b bind similarly to a minicollagen; thus, M9B collagenase molecules could initiate collagenolysis from analogous structural features in various collagen fibrils.

Some bacterial collagenases consist of up to three CBD molecules. The linker that induced structural transformation is a common feature found in M9B collagenase. It could act as a Ca^{2+} sensor to trigger domain rearrangement as a means of enzyme activation against the insoluble fibril. Physiological Ca^{2+} will likely trigger a domain rearrangement in these collagenases. Properly positioned tandem and triple CBDs should ensure that collagenase is anchored to the weakest region in fibril.

The function of the PPC domain in M9A (PPC-M9A) is unknown. Characterized M9As are collagenases, and the PPC domain from *Vibrio vulnificus* is thought to interact with type IV collagen (45). No structure of PPC-M9A has been reported. The crystal structure of PPC from serine protease S8 superfamily with a sequence identity of 15% (PDB accession code 1WME [46]) cannot offer additional insights either. The function of the PPC is not known either, but PPC-M9A may adopt a tertiary fold similar to that of CBD. Like s3, strand A in PPC-M9A may not exist. The capability to monitor the concentration change in Ca^{2+} by CBD is critical for the activation of M9B enzymes against insoluble collagen. Along with Zn^{2+} at the catalytic center, *Vibrio parahaemolyticus* M9A enzyme uses Ca^{2+} as a cofactor (47). However,

neither Ca^{2+} binding residues nor collagen binding residues identified in CBD are found in PPC-M9A. The structures of s3 and s3b could only serve as a template to the general fold of PPC-M9A.

As mentioned earlier, clinical applications using fusion proteins of CBD and various growth factors are under way. Fusion proteins of any CBD derived from M9B collagenase and a growth factor should result in comparable clinical outcomes.

ACKNOWLEDGMENTS

We thank the staff at BioCARS 14ID-B, SBC-19ID, and BM at the Advanced Photon Source, which is supported by the Department of Energy. This work has been supported by the National Institutes of Health Center for Protein Structure and Function (COBRE 8 P30GM103450 and INBRE GM103429), Arkansas Biosciences Institute, a Grant-in Aid for Scientific Research (C) from the Japan Society for the Promotion of Science, and Kagawa University Project Research Fund 2005-2006. R.B. was supported by an Honors college undergraduate research grant and a Student Undergraduate Research Fellowship from the Arkansas Department of Higher Ed.

REFERENCES

- Hurst LC, Badalamente MA, Hentz VR, Hotchkiss RN, Kaplan FT, Meals RA, Smith TM, Rodzvilla J. 2009. Injectable collagenase clostridium histolyticum for Dupuytren's contracture. *N. Engl. J. Med.* 361:968–979.
- Eckhard U, Schonauer E, Nuss D, Brandstetter H. 2011. Structure of collagenase G reveals a chew-and-digest mechanism of bacterial collagenolysis. *Nat. Struct. Mol. Biol.* 18:1109–1114.
- Wang YK, Zhao GY, Li Y, Chen XL, Xie BB, Su HN, Lv YH, He HL, Liu H, Hu J, Zhou BC, Zhang YZ. 2010. Mechanistic insight into the function of the C-terminal PKD domain of the collagenolytic serine protease *desasin MCP-01* from deep sea *Pseudoalteromonas* sp. SM9913: binding of the PKD domain to collagen results in collagen swelling but does not unwind the collagen triple helix. *J. Biol. Chem.* 285:14285–14291.
- Matsushita O, Jung CM, Minami J, Katayama S, Nishi N, Okabe A. 1998. A study of the collagen-binding domain of a 116-kDa *Clostridium histolyticum* collagenase. *J. Biol. Chem.* 273:3643–3648.
- Yeats C, Bentley S, Bateman A. 2003. New knowledge from old: in silico discovery of novel protein domains in *Streptomyces coelicolor*. *BMC Microbiol.* 3:3. doi:10.1186/1471-2180-3-3.
- Yan BQ, Chen XL, Hou XY, He H, Zhou BC, Zhang YZ. 2009. Molecular analysis of the gene encoding a cold-adapted halophilic subtilase from deep-sea psychrotolerant bacterium *Pseudoalteromonas* sp. SM9913: cloning, expression, characterization and function analysis of the C-terminal PPC domains. *Extremophiles* 13:725–733.
- Matsushita O, Koide T, Kobayashi R, Nagata K, Okabe A. 2001. Substrate recognition by the collagen-binding domain of *Clostridium histolyticum* class I collagenase. *J. Biol. Chem.* 276:8761–8770.
- Toyoshima T, Matsushita O, Minami J, Nishi N, Okabe A, Itano T. 2001. Collagen-binding domain of a *Clostridium histolyticum* collagenase exhibits a broad substrate spectrum both in vitro and in vivo. *Connect. Tissue Res.* 42:281–290.
- Wilson JJ. 2003. Ph.D. thesis. University of Arkansas, Fayetteville, AR.
- Philominathan ST, Koide T, Hamada K, Yasui H, Seifert S, Matsushita O, Sakon J. 2009. Unidirectional binding of clostridial collagenase to triple helical substrates. *J. Biol. Chem.* 284:10868–10876.
- Wilson JJ, Matsushita O, Okabe A, Sakon J. 2003. A bacterial collagen-binding domain with novel calcium-binding motif controls domain orientation. *EMBO J.* 22:1743–1752.
- Philominathan ST, Matsushita O, Gensure R, Sakon J. 2009. Ca^{2+} -induced linker transformation leads to a compact and rigid collagen-binding domain of *Clostridium histolyticum* collagenase. *FEBS J.* 276:3589–3601.
- Spiriti J, van der Vaart A. 2010. Mechanism of the calcium-induced trans-cis isomerization of a non-prolyl peptide bond in *Clostridium histolyticum* collagenase. *Biochemistry* 49:5314–5320.
- Nishi N, Matsushita O, Yuube K, Miyataka H, Okabe A, Wada F. 1998. Collagen-binding growth factors: production and characterization of functional fusion proteins having a collagen-binding domain. *Proc. Natl. Acad. Sci. U. S. A.* 95:7018–7023.

15. Katikaneni R, Ponnappakkam T, Suda H, Miyata S, Sakon J, Matsushita O, Gensure RC. 2012. Treatment for chemotherapy-induced alopecia in mice using parathyroid hormone agonists and antagonists linked to a collagen binding domain. *Int. J. Cancer* 131:E813–E821.
16. Ponnappakkam T, Katikaneni R, Miller E, Ponnappakkam A, Hirofumi S, Miyata S, Suva LJ, Sakon J, Matsushita O, Gensure RC. 2011. Monthly administration of a novel PTH-collagen binding domain fusion protein is anabolic in mice. *Calcif. Tissue Int.* 88:511–520.
17. Ponnappakkam T, Katikaneni R, Nichols T, Tobin G, Sakon J, Matsushita O, Gensure RC. 2011. Prevention of chemotherapy-induced osteoporosis by cyclophosphamide with a long-acting form of parathyroid hormone. *J. Endocrinol. Invest.* 34:e392–e397.
18. Ponnappakkam T, Katikaneni R, Suda H, Miyata S, Matsushita O, Sakon J, Gensure RC. 2012. A single injection of the anabolic bone agent, parathyroid hormone-collagen binding domain (PTH-CBD), results in sustained increases in bone mineral density for up to 12 months in normal female mice. *Calcif. Tissue Int.* 91:196–203.
19. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. 2007. Phaser crystallographic software. *J. Appl. Crystallogr.* 40:658–674.
20. Murshudov GN, Vagin AA, Dodson EJ. 1997. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* 53:240–255.
21. Minor W, Cymborowski M, Otwinowski Z, Chruszcz M. 2006. HKL-3000: the integration of data reduction and structure solution—from diffraction images to an initial model in minutes. *Acta Crystallogr. D Biol. Crystallogr.* 62:859–866.
22. Sheldrick GM. 2008. A short history of SHELX. *Acta Crystallogr. A* 64:112–122.
23. McRee DE. 1999. XtalView/Xfit—a versatile program for manipulating atomic coordinates and electron density. *J. Struct. Biol.* 125:156–165.
24. Weigand S, Stillwell B, Guise W, Quintana J, Keanne D. 2009. Flexibility and high throughput: supporting SAXS users at a joint industrial academic beamline. *Adv. X-Ray Anal.* 52:58–68.
25. Svergun D. 1992. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Cryst.* 25:495–503.
26. Svergun DI, Petoukhov MV, Koch MH. 2001. Determination of domain structure of proteins from X-ray solution scattering. *Biophys. J.* 80:2946–2953.
27. Volkov V, Svergun D. 2003. Uniqueness of ab initio shape determination in small-angle scattering. *J. Appl. Cryst.* 36:860–864.
28. Kozin M, Svergun D. 2000. A software system for rigid-body modelling of solution scattering data. *J. Appl. Crystallogr.* 33:775–777.
29. Svergun D, Barberato C, Koch M. 1995. CRY SOL—a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Crystallogr.* 28:768–773.
30. Petoukhov MV, Svergun DI. 2005. Global rigid body modeling of macromolecular complexes against small-angle scattering data. *Biophys. J.* 89:1237–1250.
31. Palma PN, Krippahl L, Wampler JE, Moura JJ. 2000. BiGGER: a new (soft) docking algorithm for predicting protein interactions. *Proteins* 39:372–384.
32. Sides CRRL, Lay JO, Philominathan ST, Matsushita O, Sakon J. 2012. Probing the 3-D structure, dynamics, and stability of bacterial collagenase collagen binding domain (apo- versus holo-) by limited proteolysis MALDI-TOF MS. *J. Am. Soc. Mass Spectrom.* 23:505–519.
33. Kirberger M, Wang X, Deng H, Yang W, Chen G, Yang JJ. 2008. Statistical analysis of structural characteristics of protein Ca²⁺-binding sites. *J. Biol. Inorg. Chem.* 13:1169–1181.
34. Philominathan ST, Koide T, Matsushita O, Sakon J. 2012. Bacterial collagen-binding domain targets undertwisted regions of collagen. *Protein Sci.* 21:1554–1565.
35. Berman HM, Westbrook JD, Gabanyi MJ, Tao W, Shah R, Kouranov A, Schwede T, Arnold K, Kiefer F, Bordoli L, Kopp J, Podvinec M, Adams PD, Carter LG, Minor W, Nair R, La Baer J. 2009. The protein structure initiative structural genomics knowledgebase. *Nucleic Acids Res.* 37:D365–D368.
36. Ding J, Choe HW, Granzin J, Saenger W. 1992. Structure of ribonuclease T1 complexed with zinc(II) at 1.8 Å resolution: a Zn²⁺ + 6H₂O carboxylate clathrate. *Acta Crystallogr. B Struct. Sci.* 48(Pt 2):185–191.
37. Zhou Y, Morais-Cabral JH, Kaufman A, MacKinnon R. 2001. Chemistry of ion coordination and hydration revealed by a K⁺ channel-Fab complex at 2.0 Å resolution. *Nature* 414:43–48.
38. Parkinson GN, Lee MP, Neidle S. 2002. Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature* 417:876–880.
39. Gros P, Kalk KH, Hol WG. 1991. Calcium binding to thermatase. Crystallographic studies of thermatase at 0, 5, and 100 mM calcium. *J. Biol. Chem.* 266:2953–2961.
40. Boel E, Brady L, Brzozowski AM, Derewenda Z, Dodson GG, Jensen VJ, Petersen SB, Swift H, Thim L, Woldike HF. 1990. Calcium binding in alpha-amylases: an X-ray diffraction study at 2.1-Å resolution of two enzymes from *Aspergillus*. *Biochemistry* 29:6244–6249.
41. Betzel C, Pal GP, Saenger W. 1988. Synchrotron X-ray data collection and restrained least-squares refinement of the crystal structure of proteinase K at 1.5 Å resolution. *Acta Crystallogr. B* 44(Pt 2):163–172.
42. Hubbard SJ, Thronton JM. 1993. NACCESS. Department of Biochemistry and Molecular Biology, University College, London, United Kingdom.
43. Perumal S, Antipova O, Orgel JP. 2008. Collagen fibril architecture, domain organization, and triple-helical conformation govern its proteolysis. *Proc. Natl. Acad. Sci. U. S. A.* 105:2824–2829.
44. Bella J. 2010. A new method for describing the helical conformation of collagen: dependence of the triple helical twist on amino acid sequence. *J. Struct. Biol.* 170:377–391.
45. Miyoshi S. 2006. *Vibrio vulnificus* infection and metalloprotease. *J. Dermatol.* 33:589–595.
46. Nonaka T, Fujihashi M, Kita A, Saeki K, Ito S, Horikoshi K, Miki K. 2004. The crystal structure of an oxidatively stable subtilisin-like alkaline serine protease, KP-43, with a C-terminal beta-barrel domain. *J. Biol. Chem.* 279:47344–47351.
47. Yu MS, Lee CY. 1999. Expression and characterization of the prtV gene encoding a collagenase from *Vibrio parahaemolyticus* in *Escherichia coli*. *Microbiology* 145(Pt 1):143–150.
48. DeLano WL. 2002. The PyMOL user's manual. DeLano Scientific, Palo Alto, CA.
49. Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22:4673–4680.